Forecasting of Billet Price using ARIMAX Model

Asmita Chotani[1]

[1]*Student, Vellore Institute of Technology, Vellore, India*

---***---

Abstract - *The Autoregressive Integrated Moving Average model of forecasting with exogenous variables (ARIMAX), has gained importance and become a popular method for prediction. The Autoregressive Integrated Moving Average model with exogenous variables is an extended version of Autoregressive Integrated Moving Average model with external variables taken into consideration. The aim of this study is to forecast the prices, production etc of raw materials that the organization utilizes. The Autoregressive Integrated Moving Average model with exogenous variables is used to forecast the price in India depending on the various factors like production of raw material and prices abroad. The model makes use of r historical data to explain the correlation between the various different factors determined in process of research and their effect on the price in India. The accuracy of the model is evaluated on the basis of mean absolute percentage error.*

Key Words: **ARIMAX, Billet Price, Steel, Prediction, Stationary Series**

1. INTRODUCTION

India being a developing country has high demands of steel**.** The market of steel in India has increased in size and so have the number of options to get the material and process them. Iron oxides are ores used for making steel. The ore is refined and then formed into pellets. The process of steel manufacturing begins with formation of coke from Coaking coal and Pellets from iron ore. The Coke and pellets are then put into a blast furnace along with limestone in order to make Pig iron. This pig iron along with scrap is then passed to a convertor. Convertor uses the iron and scrap inputs to form steel. The steel is then further refined and then casted into Billets. The iron ore fines can also be used instead of pellets for the whole process. Another way is to use an electric furnace, the only difference being that it dissolves the need of the convertor and itself does the work of both blast furnace and convertor. The Electric Furnace inputs the pellets, the Coaking coal which has undergone direct reduction and scrap to form steel which is then refined and casted. Pellets and Coke can either be formed by iron ore and Coaking coal or directly be bought form suppliers. The initial objective of the paper is to determine and visualize the various factors affecting the profit in steel manufacturing today. Further the main objective is to develop a forecasting model which could be able to predict the future prices of Billets in India. The factors to be considered in forecasting the price in India would include macroeconomic indicators, raw materials, import–export, and production of raw materials and prices of the same abroad. This paper makes use of the ARIMAX model to forecast the prices of Billet in India, taking into account the production and prices of the various mat6erials required in the creation of billets. The accuracy of the model is evaluated on the basis of Mean Absolute Percentage Error.

2. LITERATURE SURVEY

Strategic planning for the future is extremely important in any business to make definitive action plans. Due to its great importance, the knowledge of forecasting has been growing very rapidly in modern times. Forecasting has become an important aspect of all businesses. It not only has enormous economic impacts but has social and environmental impacts as well. Forecasting models help make decisions regardless of the uncertainties in future. Researchers have used different forecasting models to forecast short-term, medium term and long-term Electricity prices. The models included statistical models, Stochastic models and Multi-agent models to forecast wholesale electricity prices [1]. Further ARIMA methodology has also been used to solely forecast next-day electricity prices. Researchers created ARIMA models to forecast prices on a hourly basis [2]. Researchers have studied forecasting models to predict the temperature of the ocean and tropical circulation on a large scale to predict the climate. Similar models have been developed to provide early warnings of various storms [3]. Researchers have compared various forecasting models both qualitative and quantitative, evaluating them on the basis of maximum accuracy and minimum bias. Some of the measures used were Root Mean Squared Error, Mean Absolute Percentage error, and Mean Square Error [4]. Researchers developed forecasting models to analyze the production demand in the millennium plastic industry. The data was analyzed using double exponential smoothing and winters methods to see if the demand of products like paint buckets and dust pans were going to either decrease or increase in future demand [5]. Business activities, forecasting technologies have become indispensable tools in a wide range of administrative decision-making processes, such as finance, investments, loans, employment, banking, mortgages and investments. Banking time series data and financial time series is difficult in decomposition and forecasting due to the fact that the data has high heteroscedasticity along with being non-stationary and non-linear. Researchers have considered ARIMA as a secure and accurate model to predict Banking Stock market data for short term prediction [6]. Apart from this a lot of research papers that have been conducted in forecasting the content of stock market data. Researchers have used EMD-HW bagging method for forecasting of nonstationary and nonlinear time series data of stock price. The implementation included empirical mode decomposition (EMD), the moving block bootstrap, Holt-Winter, Intrinsic Mode Functions and Fourier Transform. Evaluating based on RMSE, MAPE, MAE, MASE and TheilU, researchers found EMW-HW bagging to be having more accuracy than traditional forecasting methods [7]. Further researchers have studied the Empirical Mode Decomposition and Moving Average Model in order to forecast stock prices for fourteen countries. Implementing, EMD, IMF and MA, the model was then compared to models like random-walk, HW and EMD without MA. Using RMSE, MAPE, MASE, MAE they concluded that EMD-MA was much more accurate than the rest of the forecasting models [8]. Further experimenting, research has been done on use of wavelet transforms and ARIMA model to forecast stock prices. Haar and Daubechies wavelet transforms were applied to the data before implementing the ARIMA model. Researchers concluded the transformed series to forecast more accurate data than normal series [9]. A combination of wavelet transforms, long-short term memory and stacked autoencoders has also been used to forecast stock prices. The wavelet transforms were used by researchers to eliminate noise in the data, SAE was used to generate high-level features which were then passed through LSTM in order to predict the stock price for the next day which is much more accurate than normal LSTM or RNN [10]. Researchers have also used the Orthogonal Wavelet Transform to smoothen the data before using it to forecast stock prices using the ARIMA model. The former method proved to be more accurate than a general ARIMA model on the basis of RMSE, MAPE and MSE measuring functions [11]. Stock prices have also been forecasted using a combination of the ARIMA model and Support Vector Machines. The use of SVM enhanced the ARIMA model and had it consider nonlinear data. The combination was found out to be more effective than the general model [12]. Researchers have also used the ARIMA model in order to forecast the production of sugarcane in Million Tons in India. It was used to forecast a single variable because it assumed and took into account the non-zero autocorrelation in between successive values of the production of sugarcane [13]. Researchers have done comparative studies on forecasting of sales in the retail industries using various methods including ARIMA, Error Trend Seasonal and Adaptive Network-based Fuzzy Inference System individually as well as in. The research concluded that combination of models gave more efficient results while use of individual models involved higher risk of inaccurate forecast[14]. ARIMA model has also been used as a mean to control congestion in LAN networks. Researchers used the ARIMA model to strategize congestion control with the data of packets circulating in a LAN network taken as the input. Further the increase in response time was forecasted and along with TCP mechanism, strategies were made [15]. Researchers have made use of ARIMA model in order to predict the air temperature and precipitation time as well. Researchers performed statistical analysis and compared them for different climate zones in order to determine the best techniques that can be made use of for the different climate zones [16]. As concluded by researchers while ARIMA is the best model for Atlantic and South Mediterranean data sets, ARIMAF models best describe the Continental precipitation. Work has also been done to forecast the demand in one of the food company's using the ARIMA model. ARIMA model has also been used in order to predict the price of Bitcoin using univariate modelling. Researchers compared the use of ARIMA model and NNAR model to predict the price of Bitcoin, and concluded that ARIMA gave more accurate results than NNAR [17]. ARIMA has also been used to forecast the prices of gold to mitigate the risk of purchasing it. Data used by the researchers was observed to have 1st order auto correlation. The researchers performed the same for 6 different ARIMA models in order to determine the best one [18]. Univariate traffic data has also been used by researchers to apply Seasonal ARIMA model on to predict the traffic flow. Researchers have worked on to make SARIMA model as one of the Intelligent Transportation Systems [19]. Along with SARIMA, researchers have also applied ARIMAX model on motorway data, which provided the researchers with improved forecasts for univariate data [20]. ARIMAX model and regression methods has been used by researchers to forecast retail sales data portraying effects of festivals as well. Further ARIMAX and regression methods have been proposed as a better and much more accurate model as compared to ARIMA and neural networks for predicting the price of retail [21]. Researchers have also used tan artificial neural network model and ARIMA model to forecast the wind speed, calculate the maximum and minimum temperatures. However, considering the RMSE, MAPE, MAE, ME and MPE the ANN model was observed to predict the wind speeds with the most accuracy [22].

3. METHODOLOGY

3.1. Introduction and Related Concepts

3.1.1.Stationarity and Differencing

Stationarity is the property of any series that does not depend on time and has constant variance. The accuracy of the data forecasted gets reduced due to the presence of trends and seasonality affecting it at random intervals of time. Due to this reason, stationarity is a required property of a series in order to forecast data.
Differencing is the process of producing a stationary series from non-stationary, it involves computing the difference between consecutive data observations (known as lags). First-order differencing treats the series for linear trends while Second-order differencing treats it for quadratic trends. In order to prevent over differencing, it is important to check if the auto correlation is not less than -0.5 and also differencing further does not increase the variance.

Stationarity Test-

Augmented Dickey Fuller (ADF) Test- A statistical significance test involving null and alternative hypothesis. With the p-value 0.05 or less, the alternative hypothesis confirms i.e. the series is stationary.

Auto-correlation Function and Partial Auto-correlation Function Plots- ACF or the auto-correlated function plot helps determine the relation between the present and past values using the differences i.e. lags taking the trends, residuals and seasonality into consideration.

On the other hand PACF takes the residuals into consideration and finds the correlation between the residuals and the next lag. Further the ACF and PACF plots help determine whether the series is stationary or not. In case of a stationary series, the ACF graph decomposes to 0 very quickly.

While PACF helps determine the order of the Autoregressive model the ACF helps determine the order of the Moving Average.

3.1.2.ARIMAX Model

Auto Regressive Integrated Moving Average (ARIMA) is a combination of both the Autoregressive model and the Moving Average model. ARIMAX is an extension of the ARIMA model, the only difference being that it considers exogenous variables as well while forecasting the data. The initial steps being the same, the exogenous variables are taken as transfer functions. In ARIMAX model the Auto regressive model makes use of the previous data while the moving average model uses the previous residuals. The model also consists of an ordinary regression model which uses external variables. Exogenous variables are basically regressors that can affect the training data or the data being forecasted.

The model is presented as ARIMA (p,d,q)

Step 1: Determine the order of differencing such that the final series is stationary. Positive autocorrelations for a large number of lags confirms that higher order of differencing. *"d"* is the order of differencing

Step 2: Determine the value of *p* and *q* using the PACF and ACF plots respectively.

Step 3: Forecast the data and calculate the mean absolute percentage error.

3.2. DATA

Data was gathered and dataset was created consisting of the information of raw materials, production, consumption, macroeconomic indicators and import/export Market research was done in order to understand the process of steel manufacturing, the processes involved, the materials required etc. to gather the information and data.

The data is mainly gathered from Steel Mint, Metal Bulletin and Ministry of Steel Websites.

In total 39 variables were identified as factors affecting the production of steel. These factors were further categorized into 5 categories as follows-

- Comparative Price- Billet Price (Global and Local)
- Raw Material- Iron Ore (Global and Local), Met Coke, Coaking Coal, Melting Scrap
- Production-Pig Iron, Pellet Sponge
- Import, Export- Steel from China, TMT from China, Semi-finished Billet, Rebar, Stainless Steel, Iron Ore, Coaking Coal, Ferrous Scrap
- Macroeconomic indicators- Rainfall, Inflation, Exchange Rate

3.3. Proposed Model

i. Data preprocessing:

The data gathered is preprocessed in order to remove duplicate data, correct errors, and remove outliers. After integrating, the data is worked upon to make it monthly data. Further units of data of same category are made the same i.e while the unit of raw materials is taken as Million Tonne (MT), price in India as INR (Rupees) and globally as USD (US dollar).

ii. Feature Engineering

Features are created with the help of the independent variables. These features were created in order to minimize the number of exogenous variables to be used as transfer function while all the information is considered.

For Eg- NET = (Production + Import) – Export

Features are basically combination of data variables which would have the same effect on the data. Another example would be Inflation and Exchange Rate, as Inflation would have a direct effect on the exchange rate.

iii. Analysis:

The data is analyzed i.e. the impact of various factors is studied. Data is analyzed in order to determine the relation between the variables and the different trends.

For eg- If the trend is same for two variables then they would have the same effect on the dependent variable, therefore any one of them can be considered for determining the forecast.
The conclusions drawn from the analysis of the various factors include-

- The import of raw materials has decreased while export has increased.
- The price of pig iron is almost the same in various cities in India
- The price of Met Coke is almost the same for CNF India and FOB China.
- The Price of Coaking Coal is almost the same for China and Australia.

iv. Modeling and Forecasting:

In most cases data till October 2019 was taken as the training data while, data from November 2019 to February 2020 were taken as the validating set. The data was tested for stationarity ad trends, further ARIMAX Models were created to forecast the price of Billet.

v. Evaluation

The Evaluation Criteria is Mean Absolute Percentage Error.

$$MAPE= \frac{\sum\left|\frac{Actual-Forecasted}{Actual}\right|}{Number\ of\ observations} * 100$$

AIC represents the quality of the developed model relatively by depicting the relative information lost in the process of creating the model with respect to the actual process used to generate the data used.

4. RESULTS AND DISCUSSIONS

Initial preprocessing of historical Billet Price Data concluded that the data had a certain amount of seasonality which made it non- stationary. The Augmented Dicker Fuller Test with the Null hypothesis being the data is non-stationary and alternative hypothesis being the data is stationary is performed. The p-value for the data differenced twice was

obtained below 0.05 making the alternative hypothesis true. Figure 1 portrays the trends and seasonality in the Training data.

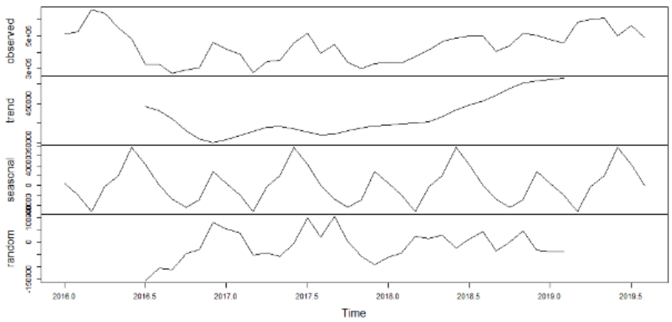

Figure 1 Decomposition of the training data series

The ACF plot portrayed in Figure 2 shows a trend in the series and thus depicts non-stationary behavior, on the other hand the ACF and PACF plots in Figure 3 portrays that the ACF value decays to zero quickly concluding that data has been made stationary.

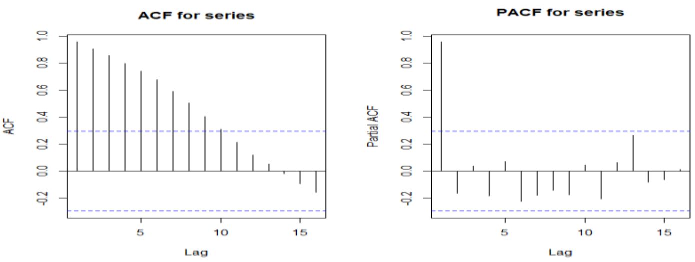

Figure 3 ACF and PACF for Training Data

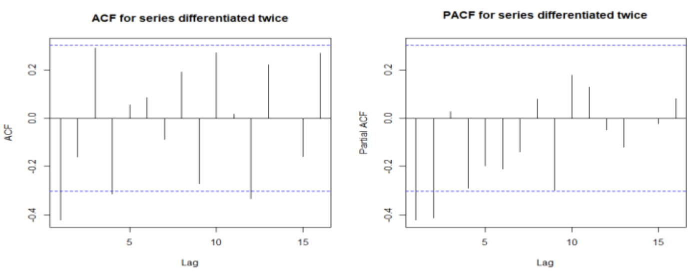

Figure 2 ACF and PACF for differentiated data

Since export has become more popular, the raw materials considered are mainly local except Coaking Coal which is imported from China. Models were developed using 14 external variables. The models with the comparitively more accuracy have been mentioned in Table1.

Evaluation of the models was done on the basis of MAPE. Further as shown in Table 1 even though the AIC value for ARIMAX(0,2,2)(1,03) is the high when compared to some of the other models, it proves be the most accurate amongst the other models in forecasting the prices due to lower

Table 1 ARIMAX MODEL RESULTS

ARIMAX Model	AIC	MAPE
(0,2,1)(1,0,3)	714.62	2.29%
(0,2,2)(1,0,3)	716.09	1.28%
(1,2,2)(1,0,3)	715.4	3.77%
(1,2,2)(1,0,2)	716.43	4.57%
(0,2,2)(1,0,2)	717.29	1.87%
(0,2,2)(2,0,2)	606.75	7.69%
(0,2,2)(2,0,3)	652.39	6.53%

5. CONCLUSION

Analysis of the Prices of Billet in India gave us ARIMAX(0,2,2)(1,0,3) and several others to predict the future prices for Billet amongst various models considered. The model considered a self-formulated dataset consisting of the production, prices, import and export of the various raw materials involved and various macroeconomic indicators. A combination of four measures was taken in order to prove the accuracy of the model. ARIMAX(0,2,2)(1,0,3) was proved to be the best model amongst the various models as it fit the statistics most appropriately and efficiently.

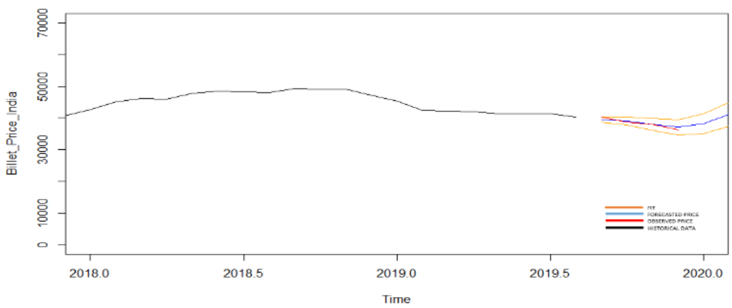

Figure 4 Forecasted Billet Price (in INR) Trend

REFERENCES

[1] Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030–1081.

[2] Contreras, J., Espinola, R., Nogales, F., & Conejo, A. (2003). ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems, 18(3), 1014–1020.

[3] J. Shukla. Predictability in the midst of chaos: A scientific basis for climate forecasting. Science, 282(5389):728–731, 1998.

[4] Wang, Shouyi & Chaovalitwongse, Wanpracha. (2011). Evaluating and Comparing Forecasting Models. Wiley Encyclopedia of Operations Research and Management Science.

[5] Daniel, E. C., Ngozi , U. M., N, M. N., & Patrick , O. U. (2014). Application of Forecasting Methods for the Estimation of Production Demand. International Journal of Science, Engineering and Technology Research (IJSETR), 3(2), 184–202.

[6] Almasarweh, M., & Wadi, S. A. (2018). ARIMA Model in Predicting Banking Stock Market Data. Modern Applied Science, 12(11), 309.

[7] Awajan AM, Ismail MT, AL Wadi S (2018) Improving forecasting accuracy for stock market data using EMD-HW bagging. PLoSONE 13(7): e0199582.https://doi.org/10.1371/journal. pone.0199582

[8] Awajan AM, Ismail MT, Al Wadi S. A hybrid emd-ma for forecasting stock market index. Italian Journal of Pure and Applied Mathematics. 2017; 38(1):1–20. 11.

[9] Al Wadia MT, Tahir Ismail M. Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model. Applied Mathematical Sciences. 2011; 5(7):315–26

[10] Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked auto encoders and long-short term memory. PloS one. 2017 Jul 14; 12(7):e0180944. https://doi.org/10.1371/journal. pone.0180944 PMID: 28708865

[11] Wadi, S. A., & Alsaraireh , A. A. (2018). INDUSTRIAL DATA FORECASTING USING DISCRETE WAVELET TRANSFORM. ITALIAN JOURNAL OF PURE AND APPLIED MATHEMATICS , 607–614.

[12] Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497-505.

[13] Kumar, Manoj & Anand, Madhu. (2014). An Application Of Time Series Arima Forecasting Model For Predicting Sugarcane Production In India. Studies in Business and Economics. 9. 81 - 94.

[14] Aras, S., Kocakoç, I. D., & Polat, C. (2017). Comparative Study On Retail Sales Forecasting Between Single And Combination Methods. Journal of Business Economics and Management, 18(5), 803–832.

[15] Sanchez, Joaquin F & Cuellar Chaves, Martha & San Mateo, Fundación Universitaria. (2016). Congestion control LAN networks using an econometric model ARIMA. Euro American Conference on Telematics and Information Systems Proceedings.

[16] Murat, M., Malinowska, I., Gos, M., & Krzyszczak, J. (2018). Forecasting daily meteorological time series using ARIMA and regression models. International Agrophysics, 32(2), 253–264.

[17] Munim, Z. H., Shakil, M. H., & Alon, I. (2019). Next-Day Bitcoin Price Forecast. Journal of Risk and Financial Management, 12(2), 103.

[18] Bandyopadhyay, G. (2016). Gold Price Forecasting Using ARIMA Model. Journal of Advanced Management Science, 117–121.

[19] Williams, Billy & Hoel, Lester. (2003). Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. Journal of Transportation Engineering. 129. 664-672.

[20] Williams, Billy. (2001). Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling. Journal of the Transportation Research Record. 1776. 194-200.

[21] Suhartono, Suhartono & Lee, Muhammad Hisyam & Prastyo, Dedy. (2015). Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects,AIP Conference Proceedings. 1691. 050026.

[22] Wesonga, Ronald & Nabugoomu, Fabian & Ababneh, Faisal & Owino, Abraham. (2018). Simulation of time series wind speed at an international airport. SIMULATION: Transactions of The Society for Modeling and Simulation International.